Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems

Evelyn Herrholz* Gerd Teschke*

November 9, 2010

Abstract

In this paper we shall be concerned with compressive sampling strategies and sparse recovery principles for linear inverse and ill-posed problems. As the main result, we provide compressed measurement models for ill-posed problems and recovery accuracy estimates for sparse approximations of the solution of the underlying inverse problem. The main ingredients are variational formulations that allow the treatment of ill-posed operator equations in the context of compressively sampled data. In particular, we rely on Tikhonov variational and constrained optimization formulations. One essential difference to the classical compressed sensing framework is the incorporation of joint sparsity measures allowing the treatment of infinite dimensional reconstruction spaces. The theoretical results are furnished with a number of numerical experiments.

Keywords: Compressive sampling, inverse and ill-posed problems, joint sparsity, sparse recovery

1 Introduction

Many applications in science and engineering require the solution of an operator equation $Kx = y$. Often only noisy data y^δ with $\|y^\delta - y\| \leq \delta$ are available, and if the problem is ill-posed, regularization methods have to be applied. During the last three decades, the theory of regularization methods for treating linear problems in a Hilbert space framework has been well developed, see, e.g., [24, 28, 29, 32, 35]. Influenced by the huge impact of sparse signal representations and the practical feasibility of advanced sparse recovery algorithms, the combination of sparse signal recovery and inverse problems emerged in the last decade

*Institute of Computational Mathematics in Science and Technology, Neubrandenburg University of Applied Sciences, Brodaer Str. 2, 17033 Neubrandenburg, Germany

as a new growing area. Currently, there exist a great variety of sparse recovery algorithms for inverse problems (linear as well as for nonlinear operator equations) within this context, see, e.g., [3, 4, 5, 15, 17, 18, 26, 27, 36, 39, 40]. These recovery algorithms are successful for many applications and have lead to breakthroughs in many fields. However, the feasibility is usually limited to problems for which the data are complete and where the problem is of moderate dimension. For really large-scale problems or problems with incomplete data, these algorithms are not well-suited or fail completely.

For the incomplete data situation, a mathematical technology, which is quite successful in sparse signal recovery, was established several years ago by D. Donoho and was called the theory of compressed sensing, see [21]. A major breakthrough was achieved when it was proven that it is possible to reconstruct a signal from very few measurements under certain conditions on the signal and the measurement model, see [8, 9, 10, 21]. First recovery results could be shown for special measurement scenarios, see [19, 20, 25], but it turned out that the theory is also applicable for more general measurement models, see e.g. [37]. The ingredients of this compressed sensing idea are as follows. Assume we are given a synthesis operator $B \in \mathbb{R}^{m\times m}$ for which a given signal $x \in \mathbb{R}^m$ has a sparse representation $x = Bd$ where d obeys just a few non-zero components. Furthermore, suppose we have a measurement matrix $A \in \mathbb{R}^{p\times m}$ which takes $p \ll m$ linear measurements of the signal x. Hence, we can describe the measuring process by $y = Ax = ABd$. A crucial property for compressed sensing to work is the so-called restricted isometry property, see [2, 10, 11, 12]. This property basically states that the product AB should have singular values either close to one (especially bounded away from zero) or zero. In [12] it was shown that if AB satisfies the restricted isometry property the solution d can be reconstructed exactly by minimization of an ℓ_1 constrained problem, provided that the solution is sparse enough. Results in [9, 22] show that even in the presence of noise, a recovery of d is possible. Up to now, all formulations of compressed sensing are finite dimensional. Quite recently, first continuous formulations have appeared for the special problem of analog-to-digital conversion, see [31, 34].

Within this paper we combine the concepts of compressive sensing and sparse recovery in inverse and ill-posed problems. To establish an adequate measurement model, we adapt an infinite dimensional compressed sensing setup that was invented in [23]. As the main result we provide recovery accuracy estimates for the computed sparse approximations of the solution of the underlying inverse problem. One essential difference to the classical compressed sensing framework is the incorporation of joint sparsity measures allowing the treatment of infinite dimensional reconstruction spaces. Moreover, we choose variational formulations that allow the treatment of ill-posed operator equations. In particular, we rely on Tikhonov variational and constrained optimization formulations.

Organization of the paper: In Section 2 we introduce the compressed measurement model and repeat some standard results in compressed sensing. In Section 3 we introduce joint sparsity measures and corresponding variational formulations and its minimization. Section 4 is devoted to the ill-posed sensing model, stabilization issues and accuracy estimates. Finally, in Section 5 we present numerical experiments.

2 Preliminaries

Within this section we provide the standard reconstruction space, the compressive sensing model and repeat classical recovery results for finite-dimensional problems that can be established thanks to the restricted isometry property of the underlying sensing matrix.

2.1 Compressive sensing model

Let X be a separable Hilbert space and $X_m \subset X$ the (possibly infinite dimensional) reconstruction space defined by

$$X_m = \left\{ x \in X,\, x = \sum_{\ell=1}^{m} \sum_{\lambda\in\Lambda} d_{\ell,\lambda} a_{\ell,\lambda}, \quad d \in (\ell_2(\Lambda))^m \right\},$$

where we assume that Λ is a countable index set and $\Phi_a = \{a_{\ell,\lambda},\, \ell = 1,\dots,m\,,\, \lambda\in\Lambda\}$ forms a frame for X_m with frame bounds $0 < C_{\Phi_a} \le C^{\Phi_a} < \infty$. Note that the reconstruction space X_m is a subspace of X with possibly large m. Typically we consider functions of the form $a_{\ell,\lambda} = a_\ell(\cdot - \lambda\mathcal{T})$, for some $\mathcal{T} > 0$. With respect to Φ_a we define the map

$$F_a \;:\; X_m \to (\ell_2(\Lambda))^m \quad\text{through}\quad x \mapsto F_a x = \begin{pmatrix} \{\langle x, a_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle x, a_{m,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} .$$

F_a is the analysis operator and its adjoint, given by

$$F_a^* \;:\; (\ell_2(\Lambda))^m \to X_m \quad\text{through}\quad d \mapsto F_a^* d = \sum_{\ell=1}^{m}\sum_{\lambda\in\Lambda} d_{\ell,\lambda} a_{\ell,\lambda} \;,$$

is the so-called synthesis operator. Since Φ_a forms a frame, we have $C_{\Phi_a} \le F_a^* F_a \le C^{\Phi_a}$ and therefore $F_a^* F_a$ is invertible implying that $I = (F_a^* F_a)^{-1} F_a^* F_a = F_a^* F_a (F_a^* F_a)^{-1}$. Consequently, each $x \in X_m$ can be reconstructed from its moments $F_a x$ through $(F_a^* F_a)^{-1} F_a^*$. A special choice of analysis/sampling functions might relax the situation a bit. Assume we have another family of sampling functions Φ_v at our disposal fulfilling $F_v F_a^* = I$, then it follows with $x = F_a^* d$

$$y = F_v x = \begin{pmatrix} \{\langle x, v_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle x, v_{m,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = \begin{pmatrix} \{\langle F_a^* d, v_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle F_a^* d, v_{m,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = F_v F_a^* d = d \;, \tag{2.1}$$

i.e. the sensed values y equal d and therefore

$$x = F_a^* F_v x \;.$$

The condition $F_v F_a^* = I$ means nothing else than $\langle a_{\ell,\lambda}, v_{\ell',\lambda'}\rangle = \delta_{\lambda\lambda'}\delta_{\ell\ell'}$ for all $\lambda, \lambda' \in \Lambda$ and $\ell, \ell' = 1,\dots,m$, i.e. Φ_v and Φ_a are biorthogonal to each other.

As we focus on reconstructing functions (or solutions of operator equations) x that have a sparse series expansion $x = F_a^* d$ with respect to Φ_a, i.e. the series expansion of x has only a very small number of non-vanishing coefficients $d_{\ell,\lambda}$, or that x is compressible (meaning that x can be well-approximated by a sparse series expansion), the theory of compressed sensing suggests to sample x at much lower rate as done in the classical setting mentioned above (there it was $m/\mathcal{T}$) while ensuring exact recovery of x (or recovery with overwhelming probability). The compressive sampling idea applied to the sensing situation (2.1) goes now as follows. Assume we are given a sensing matrix $A \in \mathbb{R}^{p\times m}$ with $p \ll m$. Then we construct p species of sampling functions through

$$\begin{pmatrix} s_{1,\lambda} \\ \vdots \\ s_{p,\lambda} \end{pmatrix} = A \begin{pmatrix} v_{1,\lambda} \\ \vdots \\ v_{m,\lambda} \end{pmatrix} \quad \text{for all } \lambda \in \Lambda \ . \tag{2.2}$$

As a simple consequence of (2.2), the following lemma holds true.

Lemma 1 *Assume for all $\lambda \in \Lambda$ the sampling functions $s_{1,\lambda}, \ldots, s_{p,\lambda}$ are chosen as in* (2.2) *and let y denote the exactly sensed data. If Φ_a and Φ_v are biorthogonal to each other, then $y = Ad$.*

Proof. Sensing x with $s_{1,\lambda}, \ldots, s_{p,\lambda}$ results in

$$y = F_s x = \begin{pmatrix} \{\langle x, s_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle x, s_{p,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = \begin{pmatrix} \{\langle F_a^* d, s_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle F_a^* d, s_{p,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = F_s F_a^* d \ .$$

By a straightforward application of (2.2) it easily follows that

$$\begin{aligned} F_s F_a^* d &= \begin{pmatrix} \{\langle F_a^* d, \sum_{i=1}^m A_{1i} v_{i,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle F_a^* d, \sum_{i=1}^m A_{pi} v_{i,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = \begin{pmatrix} \{\sum_{i=1}^m A_{1i} \langle F_a^* d, v_{i,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\sum_{i=1}^m A_{pi} \langle F_a^* d, v_{i,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} \\ &= A \begin{pmatrix} \{\langle F_a^* d, v_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle F_a^* d, v_{m,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = A F_v F_a^* d = Ad \end{aligned}$$

and the proof is complete. □

2.2 Partial recovery and classical approximation estimates

If we denote by d_λ the m-dimensional vector $(d_{1,\lambda}, \ldots, d_{m,\lambda})^T$ and by y_λ the p-dimensional vector $(y_{1,\lambda}, \ldots, y_{p,\lambda})^T$, Lemma 1 states that for each $\lambda \in \Lambda$ the measurement vectors are given by

$$y_\lambda = A d_\lambda \ . \tag{2.3}$$

It has been shown in [7], that for each individual $\lambda \in \Lambda$ the solution d_λ^* to

$$\min_{d_\lambda \in \mathbb{R}^m} \|d_\lambda\|_{\ell_1} \text{ subject to } y_\lambda = A d_\lambda \ , \tag{2.4}$$

recovers d_λ exactly provided that d_λ is sufficiently sparse and the matrix A obeys a condition known as the *restricted isometry property.*

Definition 1 (restricted isometry property) *For each integer $k = 1, 2, \ldots$, define the isometry constant δ_k of a sensing matrix A as the smallest number such that*

$$(1 - \delta_k)\|x\|_{\ell_2}^2 \leq \|Ax\|_{\ell_2}^2 \leq (1 + \delta_k)\|x\|_{\ell_2}^2 \tag{2.5}$$

holds for all k-sparse vectors x. A vector is said to be k-sparse if it has at most k non-vanishing entries.

In [6] results on the accuracy of the reconstruction from undersampled measurements are established that compare the reconstruction d_λ^* with the best k-term approximation one could obtain if the exact locations and amplitudes of the k largest entries of d_λ would be known. We denote this approximation by d_λ^k, i.e. d_λ^k is d_λ where all but the k-largest entries are set to zero.

Theorem 2 (noiseless recovery, see [6]) *Assume $\delta_{2k} < \sqrt{2} - 1$. Then for each $\lambda \in \Lambda$ the solution d_λ^* to (2.4) obeys*

$$\|d_\lambda^* - d_\lambda\|_{\ell_1} \leq C_0 \|d_\lambda^k - d_\lambda\|_{\ell_1} \tag{2.6}$$

$$\|d_\lambda^* - d_\lambda\|_{\ell_2} \leq C_0 k^{-1/2} \|d_\lambda^k - d_\lambda\|_{\ell_1} \tag{2.7}$$

for some constant C_0 (that can be explicitly computed). If d_λ is k-sparse, the recovery is exact.

This result can be extended to the more realistic scenario in which the measurements are contaminated by noise, i.e.

$$y_\lambda^\delta = A d_\lambda + z_\lambda \ , \tag{2.8}$$

where z_λ is an unknown noise term. In this setting, it is proposed in [6] to reconstruct d_λ as the solution to

$$\min_{d_\lambda \in \mathbb{R}^m} \|d_\lambda\|_{\ell_1} \text{ subject to } \|y_\lambda^\delta - A d_\lambda\|_{\ell_2} \leq \delta \ , \tag{2.9}$$

where $\|y_\lambda^\delta - y_\lambda\|_{\ell_2} \leq \delta$.

Theorem 3 (noisy recovery, see [6]) *Assume $\delta_{2k} < \sqrt{2} - 1$ and $\|y_\lambda^\delta - y_\lambda\|_{\ell_2} \leq \delta$. Then for each $\lambda \in \Lambda$ the solution d_λ^* to (2.9) obeys*

$$\|d_\lambda^* - d_\lambda\|_{\ell_2} \leq C_0 k^{-1/2} \|d_\lambda^k - d_\lambda\|_{\ell_1} + C_1 \delta \tag{2.10}$$

with the same constant C_0 as before and some C_1 (that can be explicitly computed).

3 Simultaneous recovery by joint sparsity measures

The Theorems 2 and 3 apply for all individual sensing scenarios (2.3) and (2.8), respectively (i.e. for all individual $\lambda \in \Lambda$). But as the index set Λ is possibly of infinite cardinality, we are faced with the problem of recovering infinitely many unknown vectors d_λ for which the (essential) support can be different. Therefore, the determination of d by solving for each λ an individual optimization problem is numerically not feasible.

For a simultaneous treatment of all individual optimization problems, we have to restrict the set of all possible solutions d_λ. One quite natural restriction on which we want to focus in the remaining paper is that all d_λ share a joint sparsity pattern. Introducing support sets $\mathcal{I} \subset \{1, \ldots, m\}$, a corresponding reconstruction space is given through

$$X_k = \left\{ x \in X,\, x = \sum_{\ell \in \mathcal{I}, |\mathcal{I}|=k} \sum_{\lambda \in \Lambda} d_{\ell,\lambda} a_{\ell,\lambda}, \quad d \in (\ell_2(\Lambda))^m \right\} , \tag{3.1}$$

i.e. only k out of m sequences $\{d_{\ell,\lambda}\}_{\lambda \in \Lambda}$ do not vanish. The space X_k is no longer a subspace since two different x might correspond to two different support sets $\mathcal{I}$ and therefore its sum is not contained in X_k. The space X_k can be seen as a union of (shift invariant) subspaces.

One approach to recover d was suggested in [23]. The idea goes essentially as follows: first recover the joint support $\mathcal{I}$ and reconstruct then d from given compressed samples y. The detection of $\mathcal{I}$ relies on the fact that every finite collection of vectors spanning the subspace $span(\{y_\lambda\}_{\lambda \in \Lambda})$ contains enough information to recover $\mathcal{I}$ exactly. In particular, it is possible to determine a $p \times p$ matrix V from the measurements y whose columns form a basis for the range of $\{y_\lambda\}_{\lambda \in \Lambda}$. Then, the linear system $V = AU$, in which A is the $p \times m$ compression matrix from above, has a unique k-row-sparse solution U whose support is equal to $\mathcal{I}$. Once $\mathcal{I}$ is known, system (2.3) becomes invertible. Namely, reducing A by the columns whose indices do not belong to $\mathcal{I}$ results in a matrix $A_\mathcal{I}$ for which $(A_\mathcal{I})^\dagger A_\mathcal{I} = I$, whereas $(A_\mathcal{I})^\dagger$ is the Moore-Penrose-Inverse of $A_\mathcal{I}$. Then, (2.3) can be written as $y_\lambda = A_\mathcal{I} d_{\mathcal{I},\lambda}$ and the solution can be expressed by $d_{\mathcal{I},\lambda} = (A_\mathcal{I})^\dagger y_\lambda$. It follows from the definition of the support set that all other elements in d_λ not supported on $\mathcal{I}$ are zero.

We propose an alternative by solving adequate variational problems. The essential idea to tackle the support set recovery problem is to involve a joint sparsity measure that promotes a selection of only those indices $\ell \in \{1, \ldots, m\}$ for which $\|\{d_{\ell,\lambda}\}_{\lambda \in \Lambda}\|_{\ell_r(\Lambda)}$ is large enough, i.e. where the size of the coefficients $d_{\ell,\lambda}$ indicates a significant contribution to the representation of x. We discuss two possible variational formulations which both result in iterative reconstruction procedures. The first one is the Tikhonov variational formulation for which the minimization is very easy to implement but may lack on efficiency whereas the second requires a sophisticated projection but allows an accelerated computation of an approximation of the solution. The second method even offers in a very straightforward way the involvement of ill-posed sensing models/operators.

3.1 Tikhonov variational formulation and its minimization

Let the linear sensing operator T be given by

$$T : (\ell_2(\Lambda))^m \to (\ell_2(\Lambda))^p \quad \text{via} \;\; T\,d = T \begin{pmatrix} \{d_{1,\lambda}\}_{\lambda\in\Lambda} \\ \vdots \\ \{d_{m,\lambda}\}_{\lambda\in\Lambda} \end{pmatrix} = \begin{pmatrix} \{(Ad_\lambda)^1\}_{\lambda\in\Lambda} \\ \vdots \\ \{(Ad_\lambda)^p\}_{\lambda\in\Lambda} \end{pmatrix} .$$

First we consider the Tikhonov variational formulation for the recovery of d,

$$J_\alpha(d) = \|y^\delta - Td\|^2_{(\ell_2(\Lambda))^p} + \alpha \Psi_{q,r}(d) \; . \tag{3.2}$$

The penalty term $\Psi_{q,r}$ represents the joint sparsity measure which we define for the purpose of identifying the support set $\mathcal{I}$ by

$$\Psi_{q,r}(d) = \left(\sum_{\ell=1}^{m} \left(\sum_{\lambda\in\Lambda} |d_{\ell,\lambda}|^r \right)^{\frac{q}{r}} \right)^{\frac{1}{q}} .$$

This measure forces the solution d for reasonably small chosen q (e.g. $1 \le q < 2$) to have non-vanishing rows $\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}$ only if $\|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_r(\Lambda)}$ is large enough.

Lemma 4 *The adjoint operator of T as a map between $(\ell_2(\Lambda))^p$ and $(\ell_2(\Lambda))^m$ is for each $h \in (\ell_2(\Lambda))^p$ given by*

$$T^*h = \begin{pmatrix} \{(A^Th_\lambda)^1\}_{\lambda\in\Lambda} \\ \vdots \\ \{(A^Th_\lambda)^m\}_{\lambda\in\Lambda} \end{pmatrix} .$$

Proof. This statement can be easily verified,

$$\begin{aligned} \langle Td, h\rangle_{(\ell_2(\Lambda))^p} &= \sum_{n=1}^{p}\sum_{\lambda}(Ad_\lambda)^n h^n_\lambda = \sum_{n=1}^{p}\sum_{\lambda}\left(\sum_{\ell=1}^{m} A_{n\ell}d_{\ell,\lambda}\right) h^n_\lambda \\ &= \sum_{\lambda}\langle Ad_\lambda, h_\lambda\rangle_{\mathbb{R}^p} = \sum_{\lambda}\langle d_\lambda, A^Th_\lambda\rangle_{\mathbb{R}^m} \\ &= \sum_{\ell=1}^{m}\sum_{\lambda} d_{\ell,\lambda}(A^Th_\lambda)^\ell = \langle d, T^*h\rangle_{(\ell_2(\Lambda))^m} \; . \end{aligned}$$

□

In order to compute a minimizer of (3.2) (while at the same time recovering the support of d) we apply the technique of surrogate functionals invented in [15]. Applying this technique, we obtain with $\|T\|^2 < C$ the following family of functionals

$$J^s_\alpha(d;a) = J_\alpha(d) + C\|d-a\|^2_{(\ell_2(\Lambda))^m} - \|Td - Ta\|^2_{(\ell_2(\Lambda))^p} \; , \tag{3.3}$$

which are easier to minimize and where it was shown in [15] that the sequence

$$d^{n+1} = \arg \min_{d \in (\ell_2(\Lambda))^m} J^s(d; d^n) \tag{3.4}$$

converges in norm for arbitrarily chosen d^0, $n = 0, 1, 2, \dots$, towards a minimizer of (3.2). The functional (3.3) can be written as

$$\begin{aligned} \frac{1}{C} J^s_\alpha(d; a) \quad &= \quad \|T^*(y^\delta - Ta)/C + a - d\|^2_{(\ell_2(\Lambda))^m} + \frac{\alpha}{C}\Psi_{q,r}(d) \\ &\qquad + \frac{1}{C}\|y^\delta - Ta\|^2 - \frac{1}{C^2}\|T^*(y^\delta - Ta)\|^2 \ . \end{aligned} \tag{3.5}$$

In the following proposition we provide for the case $q = 1$ the explicit description of the minimizing element d^* of (3.5).

Proposition 5 *For given a define $\tilde{a} = T^*(y^\delta - Ta)/C + a$ and let $q = 1$, then the minimizer of the surrogate functional J^s_α defined by* (3.3) *is given by*

$$\{d_{\ell,\lambda}\}_{\lambda\in\Lambda} = \left(I - P_{B_{r'}(\alpha/2C)}\right)(\{\tilde{a}_{\ell,\lambda}\}_{\lambda\in\Lambda}) \ , \quad \text{for} \quad \ell = 1, \dots, m \quad , \tag{3.6}$$

where $P_{B_{r'}(\alpha/2C)}$ denotes the orthogonal projection on the $\ell_{r'}$ ball with radius $\alpha/2C$ for which the duality relation $1/r + 1/r' = 1$ holds.

Proof. Since the second line of (3.5) does not depend on d, it remains to minimize

$$\|T^*(y^\delta - Ta)/C + a - d\|^2_{(\ell_2(\Lambda))^m} + \frac{\alpha}{C}\Psi_{q,r}(d) \ . \tag{3.7}$$

Functional (3.7) reads as

$$\begin{aligned} &\sum_{\ell=1}^m \|\{\tilde{a}_{\ell,\lambda}\}_{\lambda\in\Lambda} - \{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|^2_{\ell_2(\Lambda)} + \frac{\alpha}{C}\sum_{\ell=1}^m \|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_r(\Lambda)} \\ &\qquad = \quad \sum_{\ell=1}^m \left\{ \|\{\tilde{a}_{\ell,\lambda}\}_{\lambda\in\Lambda} - \{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|^2_{\ell_2(\Lambda)} + \frac{\alpha}{C}\|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_r(\Lambda)} \right\} \end{aligned}$$

and therefore for each $\ell = 1, \dots, m$ we have to minimize

$$\|\{\tilde{a}_{\ell,\lambda}\}_{\lambda\in\Lambda} - \{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|^2_{\ell_2(\Lambda)} + \frac{\alpha}{C}\|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_r(\Lambda)} \ .$$

We follow now a proceeding that was proposed in [17]. Consider $\|\cdot\|^*_{\ell_r(\Lambda)}$, the Fenchel transform or so–called dual functional of $\|\cdot\|_{\ell_r(\Lambda)}$, see [38]. Since $\|\cdot\|_{\ell_r(\Lambda)}$ is positive and one–homogeneous, there exists a convex set $\mathcal{C}$ such that $\|\cdot\|^*_{\ell_r(\Lambda)}$ is equal to the indicator function $\chi_{\mathcal{C}}$ over $\mathcal{C}$. In Hilbert space, we have total duality between convex sets and positive and one–homogeneous functionals, i.e. $\|\cdot\|_{\ell_r(\Lambda)} = (\chi_C)^*$, or

$$(\chi_C)^*(\cdot) = \sup_{h\in\mathcal{C}}\langle \cdot, h\rangle = \|\cdot\|_{\ell_r(\Lambda)} \ ;$$

see, e.g., [1, 13, 14]. In our case $\mathcal{C} = \{v \in \ell_2(\Lambda) : \|v\|_{\ell_{r'}(\Lambda)} \leq 1\}$ with $1/r + 1/r' = 1$. Therefore we may write

$$\|\{\tilde{a}_{\ell,\lambda}\}_{\lambda\in\Lambda} - \{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|^2_{\ell_2(\Lambda)} + \frac{\alpha}{C}\sup_{h\in\mathcal{C}}\langle\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}, h\rangle \ .$$

By arguments provided in [17, Theorem 1] we obtain

$$\{d_{\ell,\lambda}\}_{\lambda\in\Lambda} = \frac{\alpha}{2C}(I - P_{\mathcal{C}})\left(\frac{\{\tilde{a}_{\ell,\lambda}\}_{\lambda\in\Lambda}}{\alpha/2C}\right) \ ,$$

where $P_{\mathcal{C}}(v) = \arg\min_{h\in\mathcal{C}} \|h - v\|_{\ell_2(\Lambda)}$. □

In general, the computation of $I - P_{\mathcal{C}}$ is rather difficult. However, for particular choices of r' it becomes feasible, e.g. $r' = \{1, 2, \infty\}$. For the purpose of recovering the joint support with respect to $\lambda \in \Lambda$, we restrict ourselves to $r = 2$ indeed implying the desired coupling of coefficients. Setting $r = 2$ results by duality in $r' = 2$. In this situation, the projector P_{B_2} on the $\ell_2(\Lambda)$-ball of radius $\alpha > 0$ reads

$$P_{B_2(\alpha)}(x) = \begin{cases} x & , \|x\|_{\ell_2(\Lambda)} \leq \alpha \\ x\frac{1}{\|x\|_{\ell_2(\Lambda)}}\alpha & , \|x\|_{\ell_2(\Lambda)} > \alpha \end{cases} \ .$$

Consequently, we obtain

$$\begin{aligned} (I - P_{B_2(\alpha)})(x) &= \begin{cases} 0 & , \|x\|_{\ell_2(\Lambda)} - \alpha \leq 0 \\ \frac{x}{\|x\|_{\ell_2(\Lambda)}}(\|x\|_{\ell_2(\Lambda)} - \alpha) & , \|x\|_{\ell_2(\Lambda)} - \alpha > 0 \end{cases} \\ &= \frac{x}{\|x\|_{\ell_2(\Lambda)}}\max(\|x\|_{\ell_2(\Lambda)} - \alpha, 0) \qquad (3.8) \\ &=: S_\alpha(x) \ . \end{aligned}$$

The shorthand notation S_α was chosen as the resulting projector can be seen as a sequence-valued soft-shrinkage operator. Note that for $x \in \mathbb{R}$ the classical soft-shrinkage operation can be written as $\operatorname{sign}(x)\max(|x| - \alpha, 0)$. In accordance with Proposition 5, the resulting iteration can therefore be expressed for each $\ell = 1, \ldots, m$ by

$$\{d^{n+1}_{\ell,\lambda}\}_{\lambda\in\Lambda} = S_{\alpha/C}\left(\{(T^*(y^\delta - Td^n)/C)_{\ell,\lambda} + d^n_{\ell,\lambda}\}_{\lambda\in\Lambda}\right) \ . \qquad (3.9)$$

Introducing for $d \in (\ell_2(\Lambda))^m$,

$$\mathbb{S}_\alpha(d) := (\, S_\alpha(\{d_{1,\lambda}\}_{\lambda\in\Lambda}), \ldots, S_\alpha(\{d_{m,\lambda}\}_{\lambda\in\Lambda})\,) \ , \qquad (3.10)$$

a condensed notation of (3.9) is given by

$$d^{n+1} = \mathbb{S}_{\alpha/C}\left(T^*(y^\delta - Td^n)/C + d^n\right) \ . \qquad (3.11)$$

This iteration is easy to implement but, as elaborated in [16, 40], the speed of convergence is rather slow.

3.2 Constrained variational problem and its minimization

An improvement of the recovery performance can be obtained when involving the joint sparsity constraint not as an extra additive penalty term as done in (3.2) but by restricting the minimization of $\|y^\delta - Td\|^2_{(\ell_2(\Lambda))^p}$ to a reasonable sub-domain (as it was suggested and analyzed in [16, 40]). To this end, we define

$$B(\Psi_{1,2}, R) = \{d \in (\ell_2(\Lambda))^m : \Psi_{1,2}(d) \le R\}$$

and consider

$$\min_{d \in B(\Psi_{1,2},R)} \|y^\delta - Td\|^2_{(\ell_2(\Lambda))^p} \quad . \tag{3.12}$$

As $B(\Psi_{1,2}, R)$ is convex, we know from [16] that the minimizing element of $\|y^\delta - Td\|^2_{(\ell_2(\Lambda))^p}$ in $B(\Psi_{1,2}, R)$ can be approached by

$$d^{n+1} = \mathbb{P}_R\left(d^n + \frac{\gamma}{C} T^*(y^\delta - Td^n)\right) \ , \tag{3.13}$$

where $\gamma > 0$ is a step-length control (determined below) and $\mathbb{P}_R$ is the ℓ_2-projection on $B(\Psi_{1,2}, R)$. In order to verify the computational feasibility of (3.13), we show that the projection $\mathbb{P}_R$ is very easy to implement. We proceed similar as in [16] and verify first a continuity result.

Lemma 6 *For each $x \in (\ell_2(\Lambda))^m$ and all $\mu > 0$ the quantity $\Psi_{1,2}(\mathbb{S}_\mu(x))$ is piecewise linear and therefore continuous with respect to μ.*

Proof.

$$\begin{aligned}
\Psi_{1,2}(\mathbb{S}_\mu(x)) &= \sum_{\ell=1}^m \|S_\mu(\{x_{\ell,\lambda}\}_{\lambda\in\Lambda})\|_{\ell_2(\Lambda)} \\
&= \sum_{\ell=1}^m \left\| \frac{\{x_{\ell,\lambda}\}_{\lambda\in\Lambda}}{\|\{x_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}} \max(\|\{x_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \mu, 0) \right\|_{\ell_2(\Lambda)} \\
&= \sum_{\ell=1}^m \max(\|\{x_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \mu, 0) \\
&= \sum_{\|\{x_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} > \mu} (\|\{x_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \mu) \ .
\end{aligned}$$

□

The next lemma characterizes $\mathbb{P}_R$.

Lemma 7 *If $\Psi_{1,2}(x) > R$, then the ℓ_2-projection of $x \in (\ell_2(\Lambda))^m$ on $B(\Psi_{1,2}, R)$ is given by $\mathbb{P}_R(x) = \mathbb{S}_\mu(x)$, where μ as a function of x and R is chosen such that $\Psi_{1,2}(\mathbb{S}_\mu(x)) = R$. If $\Psi_{1,2}(x) \le R$, then $\mathbb{P}_R(x) = \mathbb{S}_0(x) = x$.*

Proof. Let $\Psi_{1,2}(x) > R$. By Lemma 6 there exists some $\mu = \mu(x, R) > 0$ with $\Psi_{1,2}(\mathbb{S}_\mu(x)) = R$. Due to Proposition 5 (setting $\tilde{a} = x$) and by the condensed notation (3.10), the unique minimizer of $\|x - a\|^2_{(\ell_2(\Lambda))^m} + 2\mu\Psi_{1,2}(a)$ is given by $z = \mathbb{S}_\mu(x)$, i.e.

$$\forall y \neq z : \qquad \|x - z\|^2_{(\ell_2(\Lambda))^m} + 2\mu\Psi_{1,2}(z) < \|x - y\|^2_{(\ell_2(\Lambda))^m} + 2\mu\Psi_{1,2}(y) \ .$$

Since $\Psi_{1,2}(z) = R$, it follows that

$$\forall y \in B(\Psi_{1,2}, R), y \neq z : \qquad \|x - z\|^2_{(\ell_2(\Lambda))^m} < \|x - y\|^2_{(\ell_2(\Lambda))^m} \ .$$

Consequently, z is closer to x than any other $y \in B(\Psi_{1,2}, R)$, i.e. $\mathbb{P}_R(x) = z = \mathbb{S}_\mu(x)$. □

Lemma 6 and 7 provide a simple recipe for computing the projection $\mathbb{P}_R(x)$: First, sort the $\|\cdot\|_{\ell_2(\Lambda)}$-values of the m components $\{x_{1,\lambda}\}_{\lambda\in\Lambda}, \ldots, \{x_{m,\lambda}\}_{\lambda\in\Lambda}$, resulting in a vector of sequences x^* in which the m individual rows (that possibly have infinite length) are rearranged such that $\|\{x^*_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} \geq \|\{x^*_{\ell+1,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} \geq 0$ for all $\ell = 1, \ldots, m$. Next, perform a search to find k such that

$$\begin{aligned}
\Psi_{1,2}(\mathbb{S}_{\|\{x^*_{k,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}}(x)) &= \sum_{l=1}^{k-1} \left(\|\{x^*_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \|\{x^*_{k,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}\right) \leq R \\
&< \sum_{l=1}^{k} \left(\|\{x^*_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \|\{x^*_{k+1,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}\right) \\
&= \Psi_{1,2}(\mathbb{S}_{\|\{x^*_{k+1,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}}(x)) \ .
\end{aligned}$$

Finally, set

$$\nu := k^{-1}(R - \Psi_{1,2}(\mathbb{S}_{\|\{x^*_{k,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}}(x))) \ ,$$

and

$$\mu := \|\{x^*_{k,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \nu \ .$$

With this choice we ensure that

$$\begin{aligned}
\Psi_{1,2}(\mathbb{S}_\mu(x)) &= \sum_{\ell=1}^{m} \max(\|\{x_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \mu, 0) = \sum_{\ell=1}^{k} (\|\{x^*_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \mu) \\
&= \sum_{\ell=1}^{k-1} (\|\{x^*_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \|\{x^*_{k,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}) + k\nu \\
&= \Psi_{1,2}(\mathbb{S}_{\|\{x^*_{k,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}}(x))) + k\nu = R \ .
\end{aligned}$$

Therefore, we conclude that the ℓ_2-projection on $B(\Psi_{1,2}, R)$ can be realized by the sequence-valued generalized soft-shrinkage operator. To improve the speed of convergence in (3.13) we have to specify $\gamma > 0$. To this end, by the same reasoning as in [16] or [40], we introduce conditions on γ.

Definition 2 *We say that the sequence* $\{\gamma^n\}_{n\in\mathbb{N}}$ *satisfies Condition* (B) *with respect to the sequence* $\{d^n\}_{n\in\mathbb{N}}$ *if there exists* n_0 *such that:*

$$\begin{array}{ll} \text{(B1)} & \sup\{\gamma^n; n \in \mathbb{N}\} < \infty \quad \text{and} \quad \inf\{\gamma^n; n \in \mathbb{N}\} \geq 1 \\ \text{(B2)} & \gamma^n \|Td^{n+1} - Td^n\|^2_{(\ell_2(\Lambda))^p} \leq C\|d^{n+1} - d^n\|^2_{(\ell_2(\Lambda))^m} \quad \forall n \geq n_0 \ . \end{array}$$

With the help of the condition (B) in Definition 2 on the sequence of step-lengths $\{\gamma^n\}_{n\in\mathbb{N}}$, the following convergence result can be established.

Proposition 8 *For arbitrarily chosen* d^0 *assume* d^{n+1} *is given by*

$$d^{n+1} = \mathbb{P}_R\left(d^n + \frac{\gamma^n}{C}T^*(y^\delta - Td^n)\right) \ , \tag{3.14}$$

where $\|T\|^2 < C$ *and the* γ^n *satisfy Condition* (B) *with respect to* $\{d^n\}_{n\in\mathbb{N}}$*, then the sequence of residuals* $\|y^\delta - Td^n\|^2_{(\ell_2(\Lambda))^p}$ *is monotonically decreasing. Moreover, the sequence* $\{d^n\}_{n\in\mathbb{N}}$ *converges in norm towards* d^**, where* d^* *fulfills the necessary condition for a minimum of* (3.12)*.*

4 Ill-posed sensing model and sparse recovery

Within this section we establish recovery accuracy estimates for individual vectors d^*_λ as well as for d^*. The estimates are similar to the results shown in Theorem 3, but they hold for a more general setting that also includes ill-posed compressive sensing models. In the ill-posed compressive sensing scenario, the objective is again to recover x, but we assume that we have only access to Kx, where K is supposed to be a linear (possibly ill-posed) and bounded operator between Hilbert spaces X and Y.

4.1 Sensing model

With the same analysis and synthesis operators, the data y are obtained by sensing Kx through F_s, i.e.

$$y = F_sKx = F_sKF_a^*d \ .$$

The analysis operator F_s is supposed to map between Y and $(\ell_2(\Lambda))^p$. Similarly to Lemma 1, we have the following result.

Lemma 9 *Assume for all* $\lambda \in \Lambda$ *the sampling functions* $s_{1,\lambda}, \ldots, s_{p,\lambda}$ *are chosen as in* (2.2)*. Then* $y = AF_{K^*v}F_a^*d = AF_vF_{Ka}^*d$*.*

Proof. Sensing Kx with $s_{1,\lambda}, \ldots, s_{p,\lambda}$ results in

$$y = F_sKx = \begin{pmatrix} \{\langle Kx, s_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle Kx, s_{p,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = \begin{pmatrix} \{\langle KF_a^*d, s_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle KF_a^*d, s_{p,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = F_sKF_a^*d \ .$$

Furthermore it follows,

$$\begin{aligned} F_s K F_a^* d &= \begin{pmatrix} \{\langle K F_a^* d, \sum_{i=1}^m A_{1i} v_{i,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle K F_a^* d, \sum_{i=1}^m A_{pi} v_{i,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = A \begin{pmatrix} \{\langle K F_a^* d, v_{1,\lambda}\rangle\}_{\lambda\in\Lambda} \\ \vdots \\ \{\langle K F_a^* d, v_{m,\lambda}\rangle\}_{\lambda\in\Lambda} \end{pmatrix} = A F_v K F_a^* d \\ &= A F_{K^* v} F_a^* d = A F_v F_{Ka}^* d \ , \end{aligned}$$

where the last statement is due to

$$\langle K F_a^* d, v_{\ell,\lambda}\rangle = \langle F_a^* d, K^* v_{\ell,\lambda}\rangle$$

and

$$K F_a^* d = \sum_{\ell=1}^m \sum_{\lambda\in\Lambda} d_{\ell,\lambda} K a_{\ell,\lambda}.$$

□

An ideal choice to guarantee recovery within the compressive sampling framework would be to ensure

$$F_{K^* v} F_a^* = F_v F_{Ka}^* = Id \ , \text{i.e. } \langle a_{\ell,\lambda}, K^* v_{\ell',\lambda'}\rangle = \langle K a_{\ell,\lambda}, v_{\ell',\lambda'}\rangle = \delta_{\lambda'\lambda}\delta_{\ell'\ell} \ .$$

For normalized systems Φ_a and Φ_v and ill-posed operators K this is impossible to achieve. The simplest case, which we discuss in the remaining paper, is that we have systems Φ_a and Φ_v at our disposal that diagonalize K, i.e.

$$\langle K a_{\ell,\lambda}, v_{\ell',\lambda'}\rangle = \kappa_{\ell,\lambda}\, \delta_{\lambda'\lambda}\delta_{\ell'\ell} \ . \tag{4.1}$$

One prominent example that performs such a diagonalization is the so-called wavelet-vaguelette decomposition with respect to K. If Φ_a and Φ_v diagonalize K, then the structure of the sensing operator is

$$TD : (\ell_2(\Lambda))^m \to (\ell_2(\Lambda))^p \ ,$$

where

$$(TD)\, d = (TD) \begin{pmatrix} \{d_{1,\lambda}\}_{\lambda\in\Lambda} \\ \vdots \\ \{d_{m,\lambda}\}_{\lambda\in\Lambda} \end{pmatrix} = T \begin{pmatrix} \{(D_\lambda d_\lambda)^1\}_{\lambda\in\Lambda} \\ \vdots \\ \{(D_\lambda d_\lambda)^m\}_{\lambda\in\Lambda} \end{pmatrix} = \begin{pmatrix} \{(A D_\lambda d_\lambda)^1\}_{\lambda\in\Lambda} \\ \vdots \\ \{(A D_\lambda d_\lambda)^p\}_{\lambda\in\Lambda} \end{pmatrix} \ ,$$

and D is defined by λ-dependant blocks D_λ of size $m \times m$,

$$D_\lambda = \begin{pmatrix} \kappa_{1,\lambda} & 0 & \cdots & 0 \\ 0 & \kappa_{2,\lambda} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \kappa_{m,\lambda} \end{pmatrix} \ .$$

4.2 A recovery theorem for finite dimensional subproblems

Let us consider the reconstruction problem for each individual label λ (which is an m-dimensional recovery problem). Limiting the analysis to the noisy measurement model (2.8), we have to consider

$$y_\lambda^\delta = AD_\lambda d_\lambda + z_\lambda \quad \text{with} \quad \|z_\lambda\| \leq \delta \ . \tag{4.2}$$

But since K is an ill-posed operator, we are faced with the fact that $\kappa_{\ell,\lambda}$ might become arbitrarily small. Hence, in general, the sensing matrix AD_λ obeys no longer the restricted isometry property. Therefore, we propose to reconstruct d_λ as a solution to the following stabilized constrained optimization problem

$$\min_{d_\lambda \in B(\ell_1,R)} \|y_\lambda^\delta - AD_\lambda d_\lambda\|_{\ell_2}^2 + \alpha \|d_\lambda\|_{\ell_2}^2 \ , \tag{4.3}$$

where $B(\ell_1, R) = \{d_\lambda \in \ell_2 : \|d_\lambda\|_{\ell_1} \leq R\}$ for some preassigned R. Similar combinations of ℓ_1 and ℓ_2 constraints were considered, e.g., in [30], and referred to as elastic net regularization. The difference here is the involvement of the ℓ_1 constraint requiring a different proceeding.

The minimizing element (4.3) is approximated due to the iteration (as elaborated in the previous section),

$$d_\lambda^{n+1} = \mathbb{P}_R \left(D_\lambda A^*(y_\lambda^\delta - AD_\lambda d_\lambda^n)\frac{\gamma^n}{C} + \left(1 - \frac{\alpha\gamma^n}{C}\right) d_\lambda^n \right) \ .$$

As we are interested in the accuracy of the minimizing element d_λ^* (limit of d_λ^n) of (4.3), we have to make some technical preparations.

The solution of $y_\lambda = AD_\lambda d_\lambda$ need not to be an element of $B(\ell_1, R)$. Therefore, we define the $B(\ell_1, R)$-best approximate solution as the element for which

$$d_\lambda^\dagger = \arg \min_{d_\lambda \in B(\ell_1,R)} \|y_\lambda - AD_\lambda d_\lambda\|_{\ell_2}^2$$

with

$$\|d_\lambda^\dagger\|_{\ell_2} = \min\{\|d_\lambda\|_{\ell_2} : \|y_\lambda - AD_\lambda d_\lambda\|_{\ell_2} = \|y_\lambda - AD_\lambda d_\lambda^\dagger\|_{\ell_2}\} \ .$$

Lemma 10 *Let $L^2 := D_\lambda A^* AD_\lambda + \alpha I$ and let d_λ be the solution of $y_\lambda = AD_\lambda d_\lambda$, d_λ^* the minimizer of* (4.3)*, δ the noise level as defined in* (4.2)*, and $d_\lambda^\dagger$ the $B(\ell_1, R)$-best approximate solution. Then*

$$\|L(d_\lambda^* - d_\lambda)\|_{\ell_2} \leq \sqrt{2}\|L(d_\lambda^\dagger - d_\lambda)\|_{\ell_2} + 2\delta + 3\sqrt{\alpha}R =: C(\alpha, \delta, R) \ .$$

Proof. Since d_λ^* is a minimizer of (4.3), we observe

$$\|y_\lambda^\delta - AD_\lambda d_\lambda^*\|_{\ell_2} \leq \|y_\lambda^\delta - AD_\lambda d_\lambda^\dagger\|_{\ell_2} + \sqrt{\alpha}\|d_\lambda^\dagger\|_{\ell_2} \leq \|AD_\lambda(d_\lambda^\dagger - d_\lambda)\|_{\ell_2} + \delta + \sqrt{\alpha}R \ . \tag{4.4}$$

Now we have due to the definition of L, the triangle inequality and (4.4),

$$\begin{aligned}
\|L(d_\lambda^* - d_\lambda)\|_{\ell_2} &\leq \|AD_\lambda(d_\lambda^* - d_\lambda)\|_{\ell_2} + \sqrt{\alpha}\|d_\lambda^* - d_\lambda\|_{\ell_2} \\
&\leq \|AD_\lambda d_\lambda^* - y_\lambda^\delta\|_{\ell_2} + \|AD_\lambda d_\lambda - y_\lambda^\delta\|_{\ell_2} + \sqrt{\alpha}\|d_\lambda^* - d_\lambda\|_{\ell_2} \\
&\leq \|AD_\lambda(d_\lambda^\dagger - d_\lambda)\|_{\ell_2} + \delta + \sqrt{\alpha}R + \delta + \sqrt{\alpha}\|d_\lambda^* - d_\lambda^\dagger + d_\lambda^\dagger - d_\lambda\|_{\ell_2} \\
&\leq \sqrt{2}\|L(d_\lambda^\dagger - d_\lambda)\|_{\ell_2} + 2\delta + 3\sqrt{\alpha}R \ .
\end{aligned}$$

□

The definition of L was motivated by

$$\|y_\lambda^\delta - AD_\lambda d_\lambda\|_{\ell_2}^2 + \alpha\|d_\lambda\|_{\ell_2}^2 = \|\tilde{y}_\lambda^\delta - Ld_\lambda\|_{\ell_2}^2 + \|y_\lambda^\delta\|_{\ell_2}^2 - \|L^{-1}D_\lambda A^* y_\lambda^\delta\|_{\ell_2}^2 \ , \tag{4.5}$$

where $\tilde{y}_\lambda^\delta := L^{-1}D_\lambda A^* y_\lambda^\delta$, leading for fixed $\alpha > 0$ to an optimization problem which is equivalent to (4.3), namely

$$\min_{d_\lambda \in B(\ell_1,R)} \|\tilde{y}_\lambda^\delta - Ld_\lambda\|_{\ell_2}^2 \ . \tag{4.6}$$

We have now to quantify whether optimization problem (4.6) (and (4.3), respectively) is suited to deliver a good approximation to the solution of $y_\lambda = AD_\lambda d_\lambda$, even under the presence of noise and even in the case where d_λ is not sparse (see Theorem 3 for the well-posed scenario).

If A is a feasible compression matrix, i.e. A fulfills as before the restricted isometry property (2.5), then as the basic observation, the operator L obeys

$$(\kappa_{min}^2(1-\delta_k)+\alpha)\|d_\lambda\|_{\ell_2}^2 \le \|Ld_\lambda\|_{\ell_2}^2 \le (\kappa_{max}^2(1+\delta_k)+\alpha)\|d_\lambda\|_{\ell_2}^2 \ , \tag{4.7}$$

for all k-sparse vectors d_λ and where κ_{max} denotes the largest and κ_{min} the smallest eigenvalue of D_λ. Note that in general there is no relation between the sparsity pattern of d_λ and $D_\lambda d_\lambda$. Therefore the diagonal structure of D_λ is essential.

Lemma 11 *For all $d, d' \in \mathbb{R}^m$ supported on disjoint subsets $\mathcal{I}, \mathcal{I}' \subseteq \{1, \ldots, m\}$ with $|\mathcal{I}| \le k$ and $|\mathcal{I}'| \le k'$ it holds*

$$|\langle Ld, Ld'\rangle| \le \frac{1}{2}\left\{\kappa_{max}^2 - \kappa_{min}^2 + (\kappa_{max}^2+\kappa_{min}^2)\delta_{k+k'}\right\}\|d\|_{\ell_2}\|d'\|_{\ell_2} \ . \tag{4.8}$$

Proof. As d and d' have disjoint support, we have due to (4.7),

$$(\kappa_{min}^2(1-\delta_{k+k'})+\alpha)\|d\pm d'\|_{\ell_2}^2 \le \|L(d\pm d')\|_{\ell_2}^2 \le (\kappa_{max}^2(1+\delta_{k+k'})+\alpha)\|d\pm d'\|_{\ell_2}^2 \ .$$

Assume d and d' are unit vectors, then it follows that

$$2(\kappa_{min}^2(1-\delta_{k+k'})+\alpha) \le \|L(d\pm d')\|_{\ell_2}^2 \le 2(\kappa_{max}^2(1+\delta_{k+k'})+\alpha) \ ,$$

and, moreover, by the parallelogram identity,

$$\begin{aligned} |\langle Ld, Ld'\rangle| &= \frac{1}{4}\left|\|L(d+d')\|_{\ell_2}^2 - \|L(d-d')\|_{\ell_2}^2\right| \\ &\le \frac{2}{4}\left\{\kappa_{max}^2 - \kappa_{min}^2 + (\kappa_{max}^2+\kappa_{min}^2)\delta_{k+k'}\right\} \ . \end{aligned} \tag{4.9}$$

Now, for d and d' with disjoint support (not necessarily unit length vectors) we have

$$\begin{aligned} |\langle Ld, Ld'\rangle| &= |\langle Ld/\|d\|_{\ell_2}, Ld'/\|d'\|_{\ell_2}\rangle|\,\|d\|_{\ell_2}\|d'\|_{\ell_2} \\ &\le \frac{1}{2}\left\{\kappa_{max}^2 - \kappa_{min}^2 + (\kappa_{max}^2+\kappa_{min}^2)\delta_{k+k'}\right\}\|d\|_{\ell_2}\|d'\|_{\ell_2} \end{aligned}$$

and the proof is complete. □

Theorem 12 (noisy recovery) *Assume R was chosen such that the solution d_λ does not belong to $B(\ell_1, R)$ and that*

$$0 \le \delta_{2k} < \frac{(1+\sqrt{2})\kappa_{min}^2 - \kappa_{max}^2 + \sqrt{2}\alpha}{(1+\sqrt{2})\kappa_{min}^2 + \kappa_{max}^2} \ .$$

Then the minimizer d_λ^* of (4.6) satisfies

$$\|d_\lambda^* - d_\lambda\|_{\ell_2} \le C_0 k^{-1/2} \|d_\lambda^k - d_\lambda\|_{\ell_1} + C_1 \|L(d_\lambda^\dagger - d_\lambda)\|_{\ell_2} + C_2 \delta + C_3 \sqrt{\alpha} R \ , \tag{4.10}$$

where the constants C_0, C_1, C_2, and C_3 are given explicitly.

Proof. We essentially follow the proof of Theorem 1.2 in [6]. For fixed λ set $d_\lambda^* = d_\lambda + h$ and decompose h into a sum of vectors $h_{T_0}, h_{T_1}, h_{T_2}, \ldots$, each of sparsity at most k. T_0 corresponds to the locations of the k largest coefficients of d_λ. T_1 corresponds to the locations of k largest coefficients of $h_{T_0^C}$; T_2 to the locations of the next k largest coefficients of $h_{T_0^C}$, and so on.

In a first step, we show that the size of h outside of $T_0 \cup T_1$ is essentially bounded by that of h on $T_0 \cup T_1$. In a second step, we show that $\|h_{(T_0 \cup T_1)^C}\|_{\ell_2}$ is adequately small.

First step: for each $j \ge 2$ we have,

$$\|h_{T_j}\|_{\ell_2} \le k^{1/2} \|h_{T_j}\|_{\ell_\infty} \le k^{-1/2} \|h_{T_{j-1}}\|_{\ell_1}.$$

Therefore,

$$\sum_{j \ge 2} \|h_{T_j}\|_{\ell_2} \le k^{-1/2} \left(\|h_{T_1}\|_{\ell_1} + \|h_{T_2}\|_{\ell_1} + \ldots \right) \le k^{-1/2} \|h_{T_0^C}\|_{\ell_1} \tag{4.11}$$

leading to

$$\|h_{(T_0 \cup T_1)^C}\|_{\ell_2} = \sum_{j \ge 2} \|h_{T_j}\|_{\ell_2} \le k^{-1/2} \|h_{T_0^C}\|_{\ell_1} \ . \tag{4.12}$$

As $d_\lambda \notin B(\ell_1, R)$, it follows that $\|d_\lambda^*\|_{\ell_1} < \|d_\lambda\|_{\ell_1}$. Therefore, we can verify that $\|h_{T_0^C}\|_{\ell_1}$ is reasonably bounded. We have,

$$\begin{aligned} \|d_\lambda\|_{\ell_1} \ &\ge \ \|d_\lambda^*\|_{\ell_1} = \|d_\lambda + h\|_{\ell_1} = \sum_{\ell \in T_0} |d_{\ell,\lambda} + h_\ell| + \sum_{\ell \in T_0^C} |d_{\ell,\lambda} + h_\ell| \\ &\ge \ \|(d_\lambda)_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^C}\|_{\ell_1} - \|(d_\lambda)_{T_0^C}\|_{\ell_1} \ , \end{aligned}$$

resulting in

$$\|h_{T_0^C}\|_{\ell_1} \le \|h_{T_0}\|_{\ell_1} + 2\|(d_\lambda)_{T_0^C}\|_{\ell_1} \ , \tag{4.13}$$

where by definition $\|(d_\lambda)_{T_0^C}\|_{\ell_1} = \|d_\lambda^k - d_\lambda\|_{\ell_1}$. Applying now (4.13) to (4.12) and again the Cauchy-Schwarz inequality to bound $\|h_{T_0}\|_{\ell_1}$ by $k^{1/2}\|h_{T_0}\|_{\ell_2}$, we obtain with the shorthand notation $e_0 := k^{-1/2}\|d_\lambda^k - d_\lambda\|_{\ell_1}$,

$$\|h_{(T_0 \cup T_1)^C}\|_{\ell_2} \le \|h_{T_0}\|_{\ell_2} + 2e_0 \ . \tag{4.14}$$

Second step: it remains to bound $\|h_{T_0 \cup T_1}\|_{\ell_2}$. To this end, we observe that

$$L h_{T_0 \cup T_1} = Lh - \sum_{j \ge 2} L h_{T_j} \ ,$$

and hence,

$$\|Lh_{T_0\cup T_1}\|_{\ell_2}^2 = \langle Lh_{T_0\cup T_1}, Lh\rangle - \langle Lh_{T_0\cup T_1}, \sum_{j\ge 2} Lh_{T_j}\rangle \ .$$

It follows from Lemma 10 and the restricted isometry property (4.7) for L that

$$|\langle Lh_{T_0\cup T_1}, Lh\rangle| \le \|Lh_{T_0\cup T_1}\|_{\ell_2}\|Lh\|_{\ell_2} \le C(\alpha,\delta,R)\sqrt{\kappa_{max}^2(1+\delta_{2k})+\alpha}\|h_{T_0\cup T_1}\|_{\ell_2} \ .$$

From Lemma 11 we obtain

$$|\langle Lh_{T_0}, Lh_{T_j}\rangle| \le \frac{1}{2}\left\{\kappa_{max}^2 - \kappa_{min}^2 + (\kappa_{max}^2+\kappa_{min}^2)\delta_{2k}\right\}\|h_{T_0}\|_{\ell_2}\|h_{T_j}\|_{\ell_2}$$

and

$$|\langle Lh_{T_1}, Lh_{T_j}\rangle| \le \frac{1}{2}\left\{\kappa_{max}^2 - \kappa_{min}^2 + (\kappa_{max}^2+\kappa_{min}^2)\delta_{2k}\right\}\|h_{T_1}\|_{\ell_2}\|h_{T_j}\|_{\ell_2} \ .$$

Since $\|h_{T_0}\|_{\ell_2} + \|h_{T_1}\|_{\ell_2} \le \sqrt{2}\|h_{T_0\cup T_1}\|_{\ell_2}$ for disjoint sets T_0 and T_1, it holds

$$\begin{aligned}\|Lh_{T_0\cup T_1}\|_{\ell_2}^2 &\le |\langle Lh_{T_0\cup T_1}, Lh\rangle| + \sum_{j\ge 2}\left(|\langle Lh_{T_0}, Lh_{T_j}\rangle| + |\langle Lh_{T_1}, Lh_{T_j}\rangle|\right)\\ &\le \|h_{T_0\cup T_1}\|_{\ell_2}\Bigg(C(\alpha,\delta,R)\sqrt{\kappa_{max}^2(1+\delta_{2k})+\alpha}\\ &\qquad + \frac{1}{\sqrt{2}}\left\{\kappa_{max}^2 - \kappa_{min}^2 + (\kappa_{max}^2+\kappa_{min}^2)\delta_{2k}\right\}\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2}\Bigg) \ , \qquad (4.15)\end{aligned}$$

and by left inequality of isometry property (4.7),

$$(\kappa_{min}^2(1-\delta_{2k})+\alpha)\|h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|Lh_{T_0\cup T_1}\|_{\ell_2}^2 \ . \qquad (4.16)$$

We define two auxiliary variables,

$$\theta = \frac{\sqrt{\kappa_{max}^2(1+\delta_{2k})+\alpha}}{(\kappa_{min}^2(1-\delta_{2k})+\alpha)} \ , \quad \rho = \frac{\kappa_{max}^2 - \kappa_{min}^2 + (\kappa_{max}^2+\kappa_{min}^2)\delta_{2k}}{\sqrt{2}(\kappa_{min}^2(1-\delta_{2k})+\alpha)}.$$

Dividing now (4.16) by $\kappa_{min}^2(1-\delta_{2k})+\alpha$ and bounding $\|Lh_{T_0\cup T_1}\|_{\ell_2}^2$ in (4.16) by (4.15) and applying (4.11) and (4.13) and the Cauchy-Schwarz inequality (as used after (4.13)) we obtain

$$\begin{aligned}\|h_{T_0\cup T_1}\|_{\ell_2} &\le \theta C(\alpha,\delta,R) + \rho k^{-1/2}\|h_{T_0^C}\|_{\ell_1}\\ &\le \theta C(\alpha,\delta,R) + \rho\|h_{T_0\cup T_1}\|_{\ell_2} + \rho 2e_0\end{aligned}$$

and consequently,

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le (1-\rho)^{-1}\left(\theta C(\alpha,\delta,R) + \rho 2e_0\right) \ .$$

Finally, we conclude by (4.14),

$$\begin{aligned}\|d_\lambda^* - d_\lambda\|_{\ell_2} &= \|h\|_{\ell_2} \le \|h_{T_0\cup T_1}\|_{\ell_2} + \|h_{(T_0\cup T_1)^C}\|_{\ell_2}\\ &\le 2\|h_{T_0\cup T_1}\|_{\ell_2} + 2e_0\\ &\le (1-\rho)^{-1}\left(2\theta C(\alpha,\delta,R) + (1+\rho)2e_0\right)\\ &= (1-\rho)^{-1}\left(2\theta\left\{\sqrt{2}\|L(d_\lambda^\dagger - d_\lambda)\|_{\ell_2} + 2\delta + 3\sqrt{\alpha}R\right\} + (1+\rho)2e_0\right)\end{aligned}$$

To bound the right hand side, we have to ensure that $\rho < 1$. This implies

$$(\kappa_{max}^2 + (1+\sqrt{2})\kappa_{min}^2)\delta_{2k} < (1+\sqrt{2})\kappa_{min}^2 - \kappa_{max}^2 + \sqrt{2}\alpha \ ,$$

and as $0 \leq \delta_{2k}$, we obtain the following condition on δ_{2k},

$$0 \leq \delta_{2k} < \frac{(1+\sqrt{2})\kappa_{min}^2 - \kappa_{max}^2 + \sqrt{2}\alpha}{\kappa_{max}^2 + (1+\sqrt{2})\kappa_{min}^2} \tag{4.17}$$

and the proof is complete. □

Remark 13 In the well-posed situation (see Theorem 3), perfect recovery can be achieved for $\delta = 0$ and if d_λ coincides with its best k-term approximation. In the ill-posed setting, estimate (4.10) does not provide conditions for perfect recovery. This is clear as we deal with a $B(\ell_1, R)$-best approximation framework and as the stabilization yields a changed recovery problem. Even if $\delta = 0$ and if d_λ would coincide with its best k-term approximation, the remaining terms in (4.10) do not vanish since α has to be strictly bounded away from zero and since $d_\lambda \notin B(\ell_1, R)$. If we would allow $d_\lambda \in B(\ell_1, R)$, we could not directly conclude $\|d_\lambda^*\|_{\ell_1} \leq \|d_\lambda\|_{\ell_1}$ (as needed in the proof of Theorem 12 to obtain estimate (4.13)). As d_λ^* is a minimizer of (4.6) we only have

$$\frac{\sqrt{\alpha}}{\sqrt{m}}\|d_\lambda^*\|_{\ell_1} \leq \sqrt{\alpha}\|d_\lambda^*\|_{\ell_2} \leq \|ADd_\lambda - y_\lambda^\delta\|_{\ell_2} + \sqrt{\alpha}\|d_\lambda\|_{\ell_2} = \delta + \sqrt{\alpha}\|d_\lambda\|_{\ell_2} \leq \delta + \sqrt{\alpha}\|d_\lambda\|_{\ell_1}$$

leading to $\|d_\lambda^*\|_{\ell_1} \leq \sqrt{m}\frac{\delta}{\sqrt{\alpha}} + \sqrt{m}\|d_\lambda\|_{\ell_1}$ and therefore implying another estimate than (4.13) resulting in extra constants in the final estimates of the proof.

4.3 A recovery theorem for the infinite dimensional problem

Theorem 12 provides a reasonable estimate for the reconstruction accuracy in the finite dimensional case. But practically we cannot solve infinitely many optimization problems. Therefore, we have to investigate the full infinite dimensional measurement model,

$$y^\delta = (TD)d + z \quad \text{with} \quad \|z\|_{(\ell_2(\Lambda))^m} \leq \delta \ .$$

To derive an approximation to its solution, we propose to solve the following constrained optimization problem

$$\min_{d \in B(\Psi_{1,2},R)} \|y^\delta - (TD)d\|_{(\ell_2(\Lambda))^p}^2 + \alpha\|d\|_{(\ell_2(\Lambda))^m}^2 \ . \tag{4.18}$$

Similarly as before, the minimizing element d^* is iteratively approximated by

$$d^{n+1} = \mathbb{P}_R\left(D^*T^*(y^\delta - TDd^n)\frac{\gamma^n}{C} + \left(1 - \frac{\alpha\gamma^n}{C}\right)d^n\right) \ . \tag{4.19}$$

Theorem 14 (noisy recovery) *Assume R was chosen such that the solution d of problem $y = (TD)d$ does not belong to $B(\Psi_{1,2}, R)$ and δ_{2k} is as in Theorem 12. Then the minimizer d^* of (4.18) satisfies*

$$\|d^* - d\|_{(\ell_2(\Lambda))^m} \leq C_0 k^{-1/2}\Psi_{1,2}(d^k - d) + C_1\|L(d^\dagger - d)\|_{(\ell_2(\Lambda))^m} + C_2\delta + C_3\sqrt{\alpha}R \ , \quad (4.20)$$

where the constants C_0, C_1, C_2, and C_3 are given explicitly.

Proof. We adapt the proof of Theorem 12. First we introduce the notation of row sparsity. We consider $d \in (\ell_2(\Lambda))^m$ as a semi-infinite matrix with entries $d_{\ell,\lambda}$, where $\ell = 1, \ldots, m$ is the row index and $\lambda \in \Lambda$ is the column index. Then $d \in (\ell_2(\Lambda))^m$ is said to be k-row-sparse if at most k rows are not identically zero. Furthermore, let $\mathcal{I} \subset \{1, 2, \ldots, m\}$, then $d_{\mathcal{I}}$ denotes the matrix for which all rows are set to zero except those that correspond to $\mathcal{I}$. Now set $d^* = d + h$ and decompose h into a sum of matrices $h_{T_0}, h_{T_1}, h_{T_2}, \ldots$, each of column sparsity of at most k. T_0 corresponds to the row locations of the k largest row-ℓ_2-norms $\|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}$. T_1 corresponds to the row locations of k largest row-ℓ_2-norms of $h_{T_0^C}$; T_2 to the locations of the next k largest row-ℓ_2-norms of $h_{T_0^C}$, and so on. As in the proof of Theorem 12, we proceed in two steps. First, we show that the size of h outside of $T_0 \cup T_1$ is essentially bounded by that of h on $T_0 \cup T_1$. Second, we verify that $\|h_{(T_0\cup T_1)^C}\|_{(\ell_2(\Lambda))^m}$ is adequately small.

First step: for each $j \geq 2$ we have,

$$\begin{aligned} \|h_{T_j}\|_{(\ell_2(\Lambda))^m} &\leq k^{1/2} \sup_{\ell} \|\{(h_{T_j})_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} \\ &\leq k^{-1/2} \sum_{\ell=1}^{m} \|\{(h_{T_{j-1}})_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} = k^{-1/2}\Psi_{1,2}(h_{T_{j-1}}). \end{aligned}$$

Therefore,

$$\sum_{j\geq 2} \|h_{T_j}\|_{(\ell_2(\Lambda))^m} \leq k^{-1/2}\left(\Psi_{1,2}(h_{T_1}) + \Psi_{1,2}(h_{T_2}) + \ldots\right) \leq k^{-1/2}\Psi_{1,2}(h_{T_0^C}) \quad (4.21)$$

leading to

$$\|h_{(T_0\cup T_1)^C}\|_{(\ell_2(\Lambda))^m} = \sum_{j\geq 2} \|h_{T_j}\|_{(\ell_2(\Lambda))^m} \leq k^{-1/2}\Psi_{1,2}(h_{T_0^C}) \ . \quad (4.22)$$

Since $\Psi_{1,2}(d^*) < \Psi_{1,2}(d)$, we can verify that $\Psi_{1,2}(h_{T_0^C})$ is reasonably bounded. We have,

$$\begin{aligned} \Psi_{1,2}(d) &> \Psi_{1,2}(d^*) = \Psi_{1,2}(d+h) = \sum_{\ell\in T_0} \|\{d_{\ell,\lambda} + h_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} + \sum_{\ell\in T_0^C} \|\{d_{\ell,\lambda} + h_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} \\ &\geq \sum_{\ell\in T_0} (\|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \|\{h_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}) + \sum_{\ell\in T_0^C} (\|\{h_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)} - \|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_2(\Lambda)}) \\ &\geq \Psi_{1,2}(d_{T_0}) - \Psi_{1,2}(h_{T_0}) + \Psi_{1,2}(h_{T_0^C}) - \Psi_{1,2}(d_{T_0^C}) \ , \end{aligned}$$

resulting in

$$\Psi_{1,2}(h_{T_0^C}) \leq \Psi_{1,2}(h_{T_0}) + \Psi_{1,2}(d) - \Psi_{1,2}(d_{T_0}) + \Psi_{1,2}(d_{T_0^C}) = \Psi_{1,2}(h_{T_0}) + 2\Psi_{1,2}(d_{T_0^C}) \ , \quad (4.23)$$

where by definition $\Psi_{1,2}(d_{T_0^C}) = \Psi_{1,2}(d^k - d)$. Applying now (4.23) to (4.22) and again the Cauchy-Schwarz inequality to bound $\Psi_{1,2}(h_{T_0})$ by $k^{1/2}\|h_{T_0}\|_{(\ell_2(\Lambda))^m}$, we obtain with the shorthand notation $e_0 := k^{-1/2}\Psi_{1,2}(d^k - d)$,

$$\|h_{(T_0 \cup T_1)^C}\|_{(\ell_2(\Lambda))^m} \leq \|h_{T_0}\|_{(\ell_2(\Lambda))^m} + 2e_0 \ .$$

Second step: it remains to bound $\|h_{T_0 \cup T_1}\|_{(\ell_2(\Lambda))^m}$. To this end, we have just to check whether the reasoning in the proof of Theorem 12 holds also for the $(\ell_2(\Lambda))^m$-topology. Defining $L^2 := D^*T^*TD + \alpha I$, Lemma 10 easily extends to

$$\|L(d^* - d)\|_{(\ell_2(\Lambda))^m} \leq \sqrt{2}\|L(d^\dagger - d)\|_{(\ell_2(\Lambda))^m} + 2\delta + 3\sqrt{\alpha}R =: C(\alpha, \delta, R) \ .$$

Accordingly, condition (4.7) reads as

$$(\kappa_{min}^2(1-\delta_k) + \alpha)\|d\|^2_{(\ell_2(\Lambda))^m} \leq \|Ld\|^2_{(\ell_2(\Lambda))^m} \leq (\kappa_{max}^2(1+\delta_k) + \alpha)\|d\|^2_{(\ell_2(\Lambda))^m} \ .$$

Consequently, also Lemma 11 holds true and therefore the remaining proof of Theorem 12 applies without any further changes. □

4.4 Relation between isometry constant and stabilization

As (4.17) serves as a condition for δ_{2k} and α at the same time, it turns out that the choice of α influences the choice of a suitable sensing matrix A and vice versa. Therefore, we can distinguish two scenarios: A is given in advance or can be chosen after the selection of α.

Let us first consider the scenario in which we are given a preassigned sensing matrix A and an operator K, then the choice of α is restricted to

$$\frac{(1+\delta_{2k})\kappa_{max}^2 - (1+\sqrt{2})(1-\delta_{2k})\kappa_{min}^2}{\sqrt{2}} < \alpha \ . \tag{4.24}$$

If $\delta_{2k} < 1$, a stabilization becomes necessary if

$$\frac{1+\delta_{2k}}{1-\delta_{2k}} \cdot \frac{\kappa_{max}^2}{\kappa_{min}^2} > 1 + \sqrt{2} \ . \tag{4.25}$$

If $\delta_{2k} \geq 1$, the left side in (4.24) is positive (independently on what K is) and we have to choose α accordingly. Specializing the case $\delta_{2k} < 1$ to the situation in which $\kappa_{max}^2 = \kappa_{min}^2 = 1$, inequality (4.25) simplifies to

$$\delta_{2k} > \sqrt{2} - 1 \ ,$$

reflecting a violation of the condition on δ_{2k} in Theorem 3. Condition (4.25) shows also the interplay between A and K leading to well- or ill-posedness of the problem.

Let us now consider the scenario in which we first stabilize the problem with respect to K and then select an adequate sensing matrix A. As the isometry constant δ_{2k} (not yet chosen) has to fulfill (4.17), the choice of α is limited to

$$\frac{\kappa_{max}^2 - (1+\sqrt{2})\kappa_{min}^2}{\sqrt{2}} < \alpha \ . \tag{4.26}$$

Consequently, we have to consider the case $\kappa_{max}^2 - (1+\sqrt{2})\kappa_{min}^2 \leq 0$, in which no stabilization is necessary (supposed we can choose A accordingly), and the case $\kappa_{max}^2 - (1+\sqrt{2})\kappa_{min}^2 > 0$. The first case include operators K for which

$$1 \leq \kappa_{max}^2/\kappa_{min}^2 < 1+\sqrt{2} \ . \tag{4.27}$$

By (4.17) the isometry constant has to fulfill

$$\delta_{2k} < \frac{1+\sqrt{2}-\kappa_{max}^2/\kappa_{min}^2}{1+\sqrt{2}+\kappa_{max}^2/\kappa_{min}^2} \ .$$

Due to (4.27), we observe that the bound for δ_{2k} is then allowed to vary from $\sqrt{2}-1$ (classical bound) to 0 (not including 0 itself), i.e. the larger $\kappa_{max}^2/\kappa_{min}^2$ the more must be compensated by the properties of A. This, however, might be rather difficult in practice. To this end, it might be useful to choose even in this case some $\alpha > 0$. The second case (in which we definitely must stabilize the problem) covers operators K for which

$$1+\sqrt{2} \leq \kappa_{max}^2/\kappa_{min}^2 \ . \tag{4.28}$$

As we have by (4.17),

$$\delta_{2k} < \frac{1+\sqrt{2}-\kappa_{max}^2/\kappa_{min}^2+\sqrt{2}\alpha/\kappa_{min}^2}{1+\sqrt{2}+\kappa_{max}^2/\kappa_{min}^2} \ ,$$

it follows by (4.28) that in dependance on the choice of α through condition (4.26), the bound for δ_{2k} is bounded away from 0. By setting α reasonably large, we can reduce the requirements on A. Note that in both cases choosing a large value for α may change the recovery problem significantly. This becomes visible through estimate (4.10) indicating that the recovery accuracy gets lost.

5 Numerical experiments

The following numerical experiments demonstrate the applicability of the proposed algorithms for computing sparse approximations of solutions of inverse problems on the basis of compressively sensed data. We discuss two scenarios: with and without an ill-posed operator.

Let us first consider the case which the operator $K = Id$. Then, we are only faced with a sparse signal recovery problem. In order to define Φ_a, we introduce the scaling function ϕ as the normalized cardinal sine function defined by

$$\phi(t) = \mathrm{sinc}(\pi t) = \frac{\sin(\pi t)}{\pi t} \ ,$$

which is nothing than the inverse Fourier transform of $\frac{1}{\sqrt{2\pi}}\chi_{[-\pi,\pi]}(\omega)$. All translated and dilated versions of ϕ form a multi-resolution analysis $\{V_\ell\}_{\ell\in\mathbb{Z}}$ of $X = L_2(\mathbb{R})$, see for more details [33]. The mother wavelet function a is given through $a(t) = 2\phi(2t) - \phi(t)$ with corresponding translated and dilated versions

$$a_{\ell,\lambda}(t) = 2^{\ell/2} a(2^\ell\, t - \lambda) \ ,$$

where all the translates span the detail (or complement) spaces W_ℓ that fulfil $V_{\ell+1} = V_\ell \oplus W_\ell$. This relation implies $X = V_{\ell_1} \oplus \bigoplus_{\ell=\ell_1}^{\infty} W_\ell$. To define X_m we truncate the infinite direct sum and obtain by setting $W_{\ell_0} = V_{\ell_1}$ (to simplify the notation),

$$X_m = \bigoplus_{\ell=\ell_0}^{\ell_{m-1}} W_\ell = V_{\ell_m} \ .$$

In our example we choose $\ell = -3, \ldots, 5$, i.e. $\ell_0 = -3$ and $\ell_8 = 5$ and therewith $m = 9$. As $K = Id$, we choose $\Phi_v = \Phi_a$. To determine the compressed sampling system Φ_s, we have to select a compression matrix A of size $p \times m$ (which is here a Gaussian random matrix with zero mean and variance $1/p^2$). As our synthetic example we define a 2-sparse signal and as the recovery theory requires at least $2k \leq p < m$ (k stands for the sparsity index), we pick $p = 4$. The signal x which we wish to recover is defined by

$$x(t) = 3\, a_{-3,0}(t) - 1\, a_{0,-1}(t) + 2\, a_{0,4}(t)$$

and is shown in Figure 1 (left). The right image in Figure 1 shows the corresponding coefficients $d_{\ell,\lambda}$ where only $\{d_{-3,\lambda}\}_{\lambda\in\Lambda}$ and $\{d_{0,\lambda}\}_{\lambda\in\Lambda}$ have nonzero entries, namely $d_{-3,0} = 3$, $d_{0,-1} = -1$ and $d_{0,4} = 2$. All other coefficients are set to zero. For numerical feasibility, we limit the computations to the finite interval $[-20, 20]$ which is discretized through $t_i = -20 + 0.01i$, $i = 0, 1, 2, ..., 4000$. Therefore, in the numerical simulation the index set Λ (which is allowed to be of infinite cardinality) is truncated accordingly. As we have for $K = Id$ direct access to x, we obtain our measurements y by sampling x, i.e. $y = F_s x = F_s F_a^* d = Ad$. After the sampling process, the p dimensional measurement vectors y_λ are corrupted by noise (noise level about 10 percent) resulting in y^δ. Now we apply the proposed iterations (3.11) and (3.14) (with and without acceleration) to recover d and therewith x. As the initial guess for all iterations we have always chosen d^0 identically zero. Figure 2 shows the recovery result for (3.11). The constant C in (3.11) must fulfill $\|T\|^2 < C$. Since $\|T\| \leq \|A\|$, the constant C must be an upper bound for the maximal eigenvalue of A^*A. The recovery results in Figure 2 are obtained for $\alpha = 0.5$ and after 150 iteration steps. As an observation

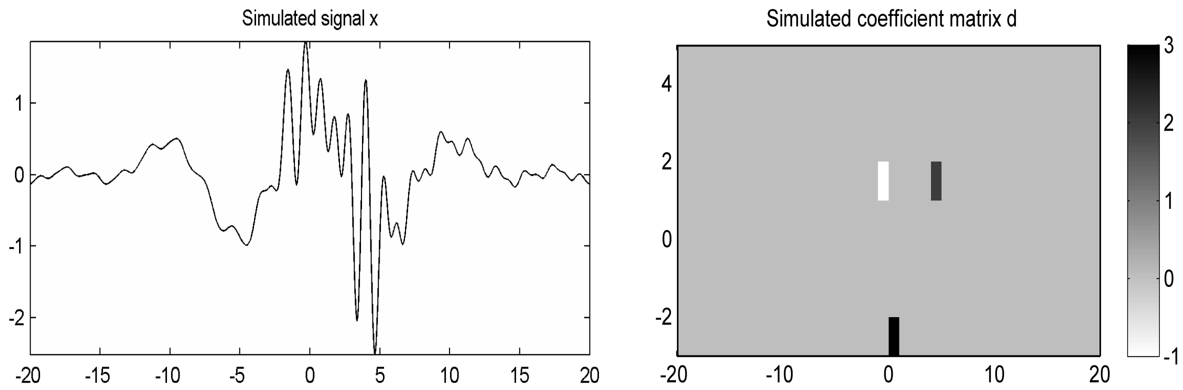

Figure 1: Simulated signal x (left) and the corresponding coefficients d (right). In this case, x is 2-sparse meaning that two out of $m = 9$ sequences $\{d_{\ell,\lambda}\}_{\lambda \in \Lambda}$ have nonzero elements.

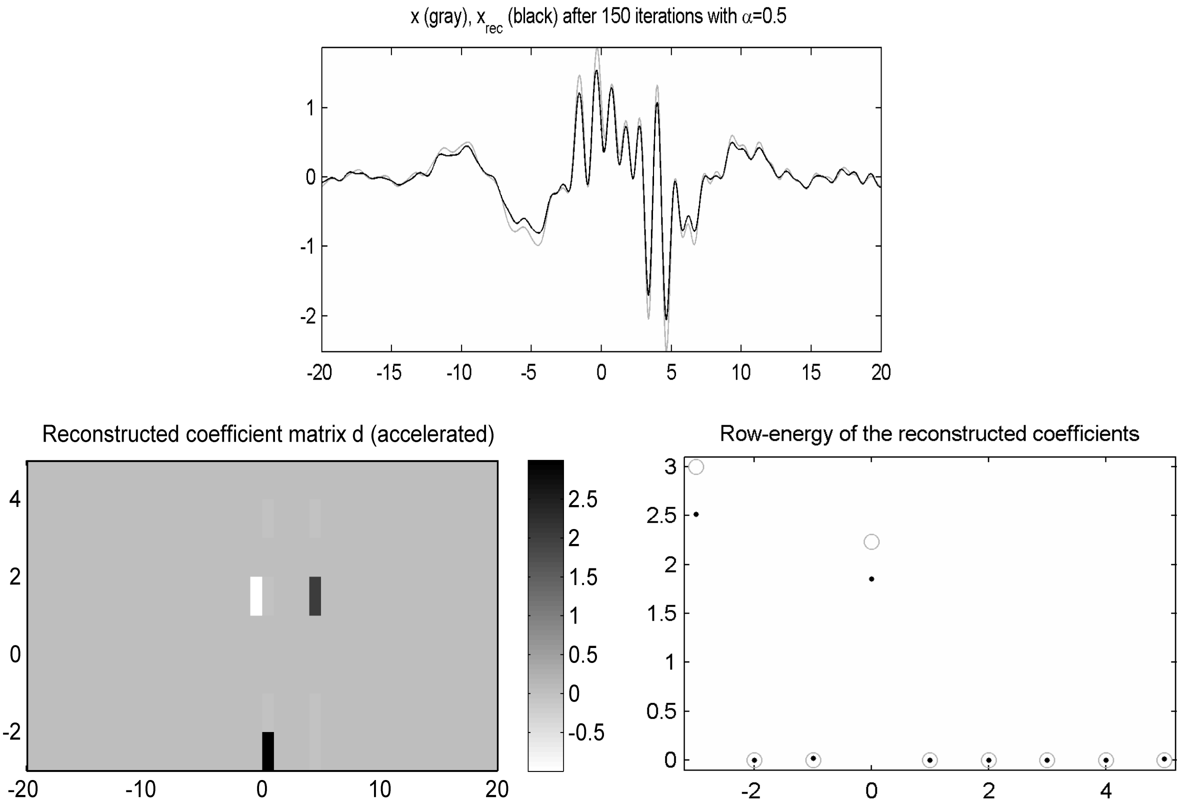

Figure 2: Recovery signal obtained with method (3.11). Top image: recovered function x^*, bottom left: recovered coefficients d^*, bottom right: row ℓ_2-energy of d^*.

(which was still recognized and extensively discussed in [16]), this method converges quite slow. Nevertheless, the joint support of d is correctly identified (compare with the bottom right image in Figure 2). A significant improvement is obtained with iteration (3.14). The bottleneck of this procedure is the choice of R. This, of course, requires a-priori knowledge of the solution to be reconstructed. In our synthetic experiment we pick $R = \sqrt{5} + \sqrt{9}$, which

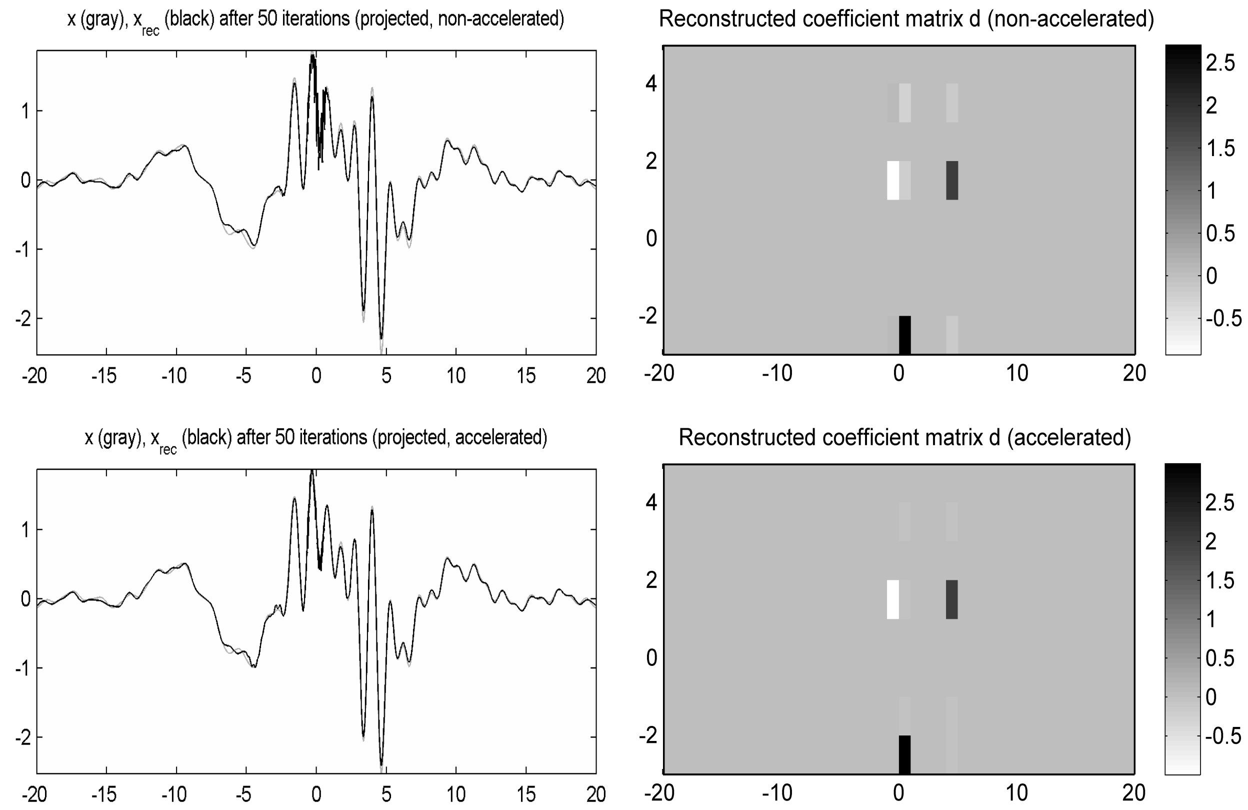

Figure 3: Top row: reconstruction of x (left) obtained with the projected iteration (3.14) with constant step length (non-accelerated iteration) and corresponding coefficients d^*, bottom row: reconstruction of x (left) obtained with the projected iteration (3.14) with variable step length (accelerated iteration) and corresponding coefficients d^*.

is equal to $\Psi_{1,2}(d)$. In the first experiment we set $\gamma = 1$, which is an iteration with constant step length. The recovery results of this method after 50 iteration steps are illustrated in the top row in Figure 3. The reconstructed signal is displayed on the left and the coefficients on the right side. The speed of convergence of the projected method can be significantly improved by allowing γ to vary in accordance with Condition (B). The values for γ are determined during the iteration leading much better convergence properties of (3.14). The recovery results are shown in the bottom row in Figure 3. The differences of both schemes can be observed in Figure 4. The iterates of both the projected method (gray dots) and the accelerated projected method (black dots) live (from a certain number of iterations on) on the boundary of $B(\Psi_{1,2}, R)$ which can be seen in Figure 4 (left). Furthermore, the effect of the acceleration can be observed in Figure 4 (right). The residual value obtained by the non-accelerated version after 50 iteration steps is already achieved by the accelerated variant after 5 steps.

In the second experiment we discuss the case in which K is a non-trivial operator. We

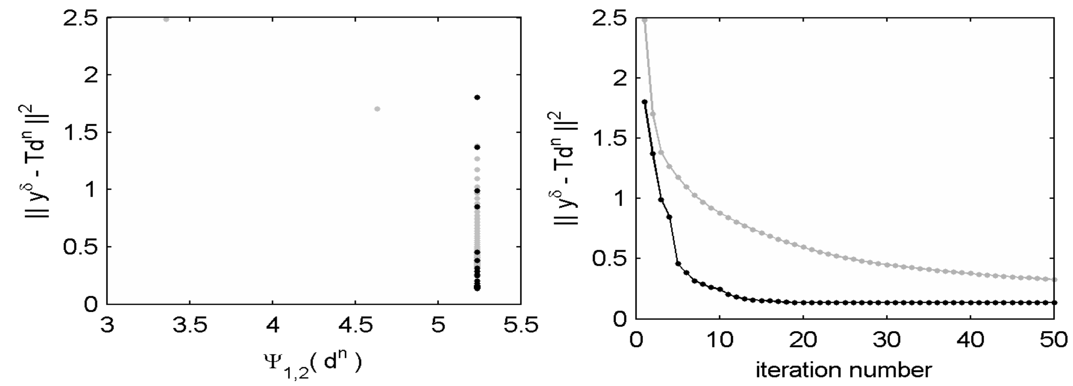

Figure 4: Left image: sparsity to residual plot of the accelerated (black) and non-accelerated (gray) projected iteration, right image: decay of residuum with respect to the number of iterations for the accelerated (black) and non-accelerated (gray) scheme.

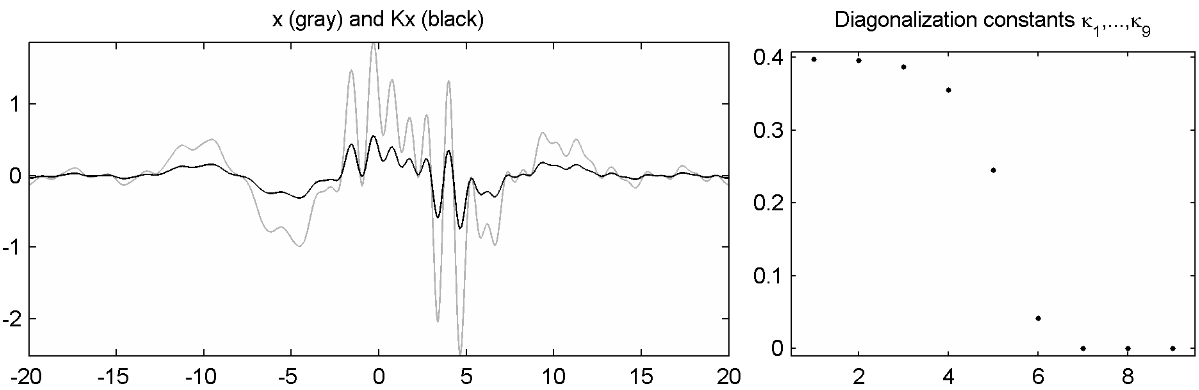

Figure 5: Left: original signal x (gray) and the convolution of x with the Gaussian (black), right: diagonalization constants κ_ℓ.

focus on an inverse problem which is given by an integral operator equation $Kx = y$,

$$y(t) = Kx(t) = \int g(t-\tau)x(\tau)d\tau \ ,$$

where g is the Gaussian function (with zero mean and variance 0.01) and $K : X \to Y$. If we deal with noisy data y^δ, we always assume $\|y - y^\delta\| \leq \delta$. As mentioned in Section 2.1, we restrict ourselves to the case in which the solution of the inverse problem belongs to $X_m \subset X$ which is spanned by Φ_a. In this experiment Φ_a again consists of sinc-type wavelet functions. This choice is practically motivated as the resulting function space is a space of bandlimited functions. Moreover, the choice of a wavelet system allows a straightforward diagonalization of K due to the corresponding wavelet-vaguelette-decomposition of K, see [33]. The system Φ_v, which performs the analysis/sensing step, then consists of the associated collection of vaguelette functions. The vaguelette construction principles can be retraced in [33]. Assume K is a bounded and linear operator and Φ_a is defined as above, then the vaguelette functions

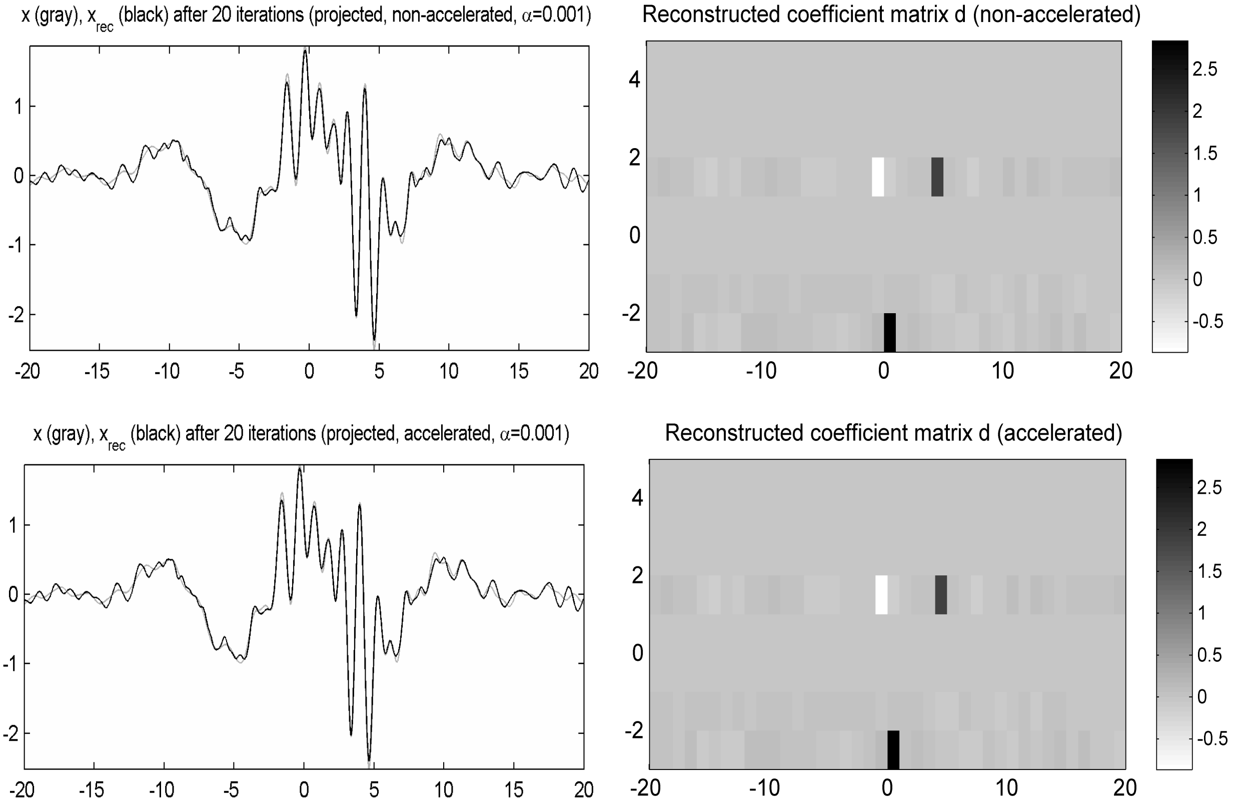

Figure 6: Top row: reconstruction of x (left) obtained with the projected iteration (4.19) with constant step length (non-accelerated iteration) and corresponding coefficients d^*, bottom row: reconstruction of x (left) obtained with the projected iteration (4.19) with variable step length (accelerated iteration) and corresponding coefficients d^*.

$v_{\ell,\lambda}$ are defined through

$$K^* v_{\ell,\lambda} = \kappa_{\ell,\lambda} a_{\ell,\lambda} \quad \text{with} \quad \|v_{\ell,\lambda}\|_Y = 1 \quad \forall \ell, \lambda \ .$$

If the system Φ_v is constructed as mentioned above, we obviously have $\langle K a_{\ell,\lambda}, v_{\ell',\lambda'} \rangle = \kappa_{\ell,\lambda} \delta_{\lambda'\lambda} \delta_{\ell'\ell}$. In the particular case of convolution operators the numbers $\kappa_{\ell,\lambda}$ just depend on ℓ, i.e. $\kappa_{\ell,\lambda} = \kappa_\ell$ for all $\lambda \in \Lambda$. The number κ_ℓ in our specific example are visualized in Figure 5 (right). The compressed analysis/sensing system Φ_s is generated as before by (2.2) leading to the analysis operator F_s. Sensing now Kx yields, as shown in Lemma 9,

$$y = F_s K x = A F_{K^* v} F_a^* d = T D d \ .$$

In Figure 5 the original signal x and its smoothed version Kx is illustrated. The data y are contaminated by additive noise (again 10 percent). The recovery of d is performed by iteration (4.19). The difference to (3.14) is the extra parameter α that realizes the stabilization of the problem and is chosen here $\alpha = 0.00003$. The results of both, the non-accelerated and the accelerated variant are shown in Figure 6. Notice that here just 20 iteration steps were done. The reconstruction of x is rather good.

Summarizing our numerical findings, we can conclude that for the well-posed as well as for the ill-posed scenario the formulated optimization problems involving joint sparsity constraints seem to be well-suited. This can be verified due to the fact that the recovery the joint support of d and finally the recovery of a sparse approximation of d could be done quite effectively. The choice of α in accordance with Section 4.4 has required a good estimate of δ_{2k}. This was done experimentally. The choice of R, however, has requested a-priori knowledge on the sparsity of the signal to recovered. If now no knowledge on the sparsity is available, there exists an heuristic proceeding that was suggested in [16] and goes as follows: choose a slowly increasing radius, i.e.

$$R^n = (n+1)R/N \ ,$$

where n is the iteration index and N stands for a prescribed number of iterations. This proceeding yields in all considered experiments reasonable results. If R is chosen too large it may easily happen that the ℓ_1 constraint has almost no impact and the solution can be arbitrarily far off the true solution. However, convergence of a scheme with varying R^n is theoretically not verified yet.

References

[1] G. Aubert and J.-F. Aujol. Modeling very oscillating signals. Applications to image processing. *INRIA report No 4878, ISSN 0249-6399*, 2003.

[2] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. Constructive Approximation, 28(3):253–263, 2008.

[3] Thomas Bonesky, Kristian Bredies, Dirk A. Lorenz, and Peter Maass. A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. *Inverse Problems*, 23:2041–2058, 2007.

[4] K. Bredies, D.A. Lorenz, and P. Maass. A generalized conditional gradient method and its connection to an iterative shrinkage method. *Computational Optimization and Application*, 42(2):173–193, 2009.

[5] Kristian Bredies and Dirk A. Lorenz. Iterated hard shrinkage for minimization problems with sparsity constraints. *SIAM Journal on Scientific Computing*, 3(2):215–232, 2008.

[6] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l'Academie des Sciences*, "Paris, Serie I, 346":589–592, 2008.

[7] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[8] Emmanuel Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52(2):489–509, 2006.

[9] Emmanuel Candès, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[10] Emmanuel Candès and Terence Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Information Theory*, 52(12):5406–5425, 2006.

[11] Emmanuel J. Candès. Compressive sampling. In *Proc. International Congress of Mathematics*, pages 1433–1452, 2006.

[12] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Transaction on Information Theory*, 51(12):4203–4215, 2005.

[13] A. Chambolle. An Algorithm for total variation minimization and applications. *CEREMADE - CNRS UMR 7534.*

[14] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward–backward splitting. *Multiscale Modeling and Simulation*, 4(4), 2005.

[15] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[16] I. Daubechies, M. Fornasier, and I. Loris. Accelerated projected gradient methods for linear inverse problems with sparsity constraints. *J. Fourier Anal. Appl.*, 14(5-6):764–792, 2008.

[17] I. Daubechies, G. Teschke, and L. Vese. Iteratively solving linear inverse problems with general convex constraints. *Inverse Problems and Imaging*, 1(1):29–46, 2007.

[18] I. Daubechies, G. Teschke, and L. Vese. On some iterative concepts for image restoration. *Advances in Imaging and Electron Physics*, 150:1–51, 2008.

[19] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proc. Natl. Acad. Sci.*, 100:2197–2202, 2003.

[20] D. L. Donoho and X. Hou. Uncertainty priciples and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47:2845–2862, 2001.

[21] David Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[22] David L. Donoho, Michael Elad, and Vladimir Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52:6–18, 2006.

[23] Y. C. Eldar. Compressed Sensing of Analog Signals in Shift-Invariant Spaces. *IEEE Trans. on Signal Processing*, 57(8), 2009.

[24] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[25] A. Feuer and A. Nemirovski. On sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49:1579–1581, 2003.

[26] M. Fornasier. Domain decomposition methods for linear inverse problems with sparsity constraints. *Inverse Problems*, 23:2505, 2007.

[27] M. Fornasier and H. Rauhut. Recovery algorithms for vector valued data with joint sparsity constraint. *SIAM J. Numer. Anal.*, 46(2):577–613, 2008.

[28] C. W. Groetsch. *The Theory of Tikhonov regularization for Fredholm Equations of the First Kind*. Pitman, Boston, 1984.

[29] C. W. Groetsch. *Inverse Problems in the Mathematical Sciences*. Vieweg, Braunschweig, 1993.

[30] B. Jin, D. A. Lorenz, and S. Schiffler. Elastic-Net Regularization: Error estimates and Active Set Methods. *Inverse Problems*, 25(11).

[31] J. Laska, S.Kirolos, M. Duarte, T. Ragheb, R. Baraniuk, and Y. Massoud. Theory and implementation of an analog-to-information converter using random demodulation. In *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS), New Orleans*, 2007.

[32] A. K. Louis. *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.

[33] A. K. Louis, P. Maaß, and A. Rieder. *Wavelets*. Teubner, Stuttgart, 1998.

[34] M. Mishali and Y. C. Eldar. Blind multi-band signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. on Signal Processing*, 57(3):993–1009, 2009.

[35] V. A. Morozov. *Methods for Solving Incorrectly Posed Problems*. Springer, New York, 1984.

[36] R. Ramlau and G. Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numerische Mathematik*, 104(2):177 – 203, 2006.

[37] Holger Rauhut, Karin Schass, and Pierre Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Trans. Inform. Theory*, 54(5):2210–2219, 2008.

[38] R. T. Rockafellar and R. J-B. Wets. *Variational Analysis.* Springer, Berlin, 1998.

[39] G. Teschke. Multi-frame representations in linear inverse problems with mixed multi-constraints. *Appl. Computat. Harmon. Anal.*, 22:43 – 60, 2007.

[40] G. Teschke and C. Borries. Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints. *Inverse Problems*, 26:025007 (doi:10.1088/0266–5611/26/2/025007), 2010.